EOSDIS Evolution at the NASA LaRC Atmospheric Science Data Center (ASDC)

CERES Science Team Meeting

May 3, 2006

Michelle T. Ferebee
Head, ASDC
Michelle.T.Ferebee@nasa.gov

Agenda

- ASDC Organizational Changes
- Current Environment
- Overview of EOSDIS Evolution
- ASDC Evolution Details
- Current Status of ASDC/CERES Evolution
- Next Steps

Organization Chart

E3

Science Directorate
Director: Dr. Lelia Vann
Deputy: Dr. Malcolm Ko (Acting)
Lead Scientist: Dr. Malcolm Ko
Secretary: Angie Davis
Business Manager: Dee Poupard
Analyst: Garnett Hutchinson

DEVELOP
National Director: Mike Ruiz

Science Advisory Board

Senior Scientist for Earth: Dr. Bruce Wielicki
Senior Scientist for Space: Dr. Jay Bergstahl
Outreach Manager: Dr. Waldo Rodriguez
Applications Manager: Dr. Richard Eckman
IT Manager: John Kusterer
Technology Manager: Dr.Brian Killough
New Business Manager: Wallace Harrison
Facility Coordinator: Tonya Grigsby
Safety Head: Tonya Grigsby

Performs function for organization but physically located within one of the research branches

Atmospheric Science Data Center
Head: Michelle Ferebee
Secretary: Joella Hanlon
Systems Engineer: Dave Cordner
IT Research Manager: Juliet Pao
Computer Engineer: Erica Alston

E301

Boxes may change as business changes

Climate Science Branch
Head: Gary Gibson
Asst: Dr. Kory Priestley
Administrative Assistant: Shannon Walker
Scientists/Researchers
Computer Scientists
Technician

E302

Chemistry & Dynamics Branch
Head: Dave Young
Asst.: Dr. Richard Eckman
Administrative Assistant: Jeanette Reavis
Scientists/Researchers
Technicians

E303

ASDC Background

- ASDC is a full service data center for the **production**, **archival**, and **distribution** of Earth Science data
- Supports the science disciplines of **Radiation Budget, Aerosols, Clouds, and Tropospheric Chemistry**
- Operations began in January 1993
  - EOSDIS Version 0 Interoperability Team received Golden Hammer Award
- Data is distributed to an international user community
  - 15,262 customers from more than 150 countries received 125 TB of data in FY05
- Currently supporting over 40 science projects (teams) with over 1000 data sets in the archival and distribution system
  - Archive Growth Rate: 21 TB/month
    - Ingest Rate: 5 TB/month
    - Production Rate: 16 TB/month
- ASDC has 2 Science Data Processing Systems
  - LATIS: Developed in-house
  - EOSDIS Core System (ECS): Developed by Raytheon

ASDC Services

- Design, develop, and operate advanced science data production, archival, and distribution systems
  – Augmented LATIS to increase processing rates from 3X to 10X for key subsystems
  – Utilized Simple, Scalable, Script-based Science Processor Missions (S4PM) framework from GSFC for automated systems (e.g. CALIPSO Automated Processing System (CAPS), MISR S4PM, SRB S4P, FLASHFlux S4PM systems)
- Provide 7-day per week operations to support satellite missions and field campaigns
- Create web-based information management systems for science missions and for general public use
  – GEWEX RFA site
- Perform "value-added" data activities: Subsetting, Visualization, Data Conversion and Re-gridding
- Provide user assistance on science data sets, metadata, data formats, documentation, read software, and data systems

Drivers for Evolution

- Recurring budget pressures caused need to identify how to reduce operations costs while maintaining services
- Aging Hardware
- ESDIS encouraged DAACs to “think outside the box” to improve our efficiency and services

EOSDIS Evolution

- Earth Science Data and Information System (ESDIS) at NASA GSFC embarking on an EOSDIS Evolution process
  - Began in Feb 05 with goals set by HQ appointed Study Team
  - Plan based on 13 informal responses received from EOSDIS Elements (DAACs, SIPS, ECS)
  - 4 responders were selected to proceed (MODAPS, ECS, GES DAAC, ASDC)
- Changes designed to
  - Improve efficiency and introduce more autonomy, agility and scalability through infusion of newer commodity based hardware
  - Move control over processing, archive and distribution for specific science data to science teams
  - Reduce annual operational costs by 15-25% within 3 years
- Changes will happen gradually over next 2-3 years and are planned to
  - Reduce risk associated with operational changes
  - Have “proof of value” periods before taking next steps
  - Continue full operations while evolving

MODAPS

Transfer responsibility for MODIS processing, archive and distribution from GES DAAC to MODAPS

Initial Step: Atmosphere data

Features:

- Processing on demand for Level 1
- All commodity-based for processing, on-line data storage and access
- Based on extension to existing MODAPS science team production systems

Benefits:

- Reduction in archive growth through on-demand processing
- Faster access to products, reduced reprocessing time from all on-line storage
- Reduced costs due to use of commodity disks, reduction in operations at DAAC
  - Resulting changes to GES DAAC include: 10% fewer products; 90% reduction in total volume and similar reductions in archive growth/day
- Closer involvement and control by science community

Future Steps (contingent on success of Initial Step) include transition of Land data and Snow & Ice data in coordination with HQ Measurement Systems planning

GES DAAC

Consolidate GES DAAC data holdings into one DAAC-unique system (S4PA)

Features:

- Transition of AIRS, HIRDLS, OMI and V0/V1 managed heritage data sets
- Phase out of ECS in FY08 timeframe
- Based on existing on-going effort (development already in progress)
- Reduction of archive volume (due to transfer of MODIS data)

Benefits:

- Reduction in operations costs due to elimination of multiple systems
- Reduction in sustaining engineering costs due to use of simpler, scalable software and reduction in dependency on COTS products
- Increased system automation due to single system, simpler operational scenarios
- Improved data access due to planned use of increased on-line storage and commodity disks/platforms (from ECS reduced footprint plan)
- Phased elimination provides risk mitigation for MODAPS effort

ECS/SDPS

Rearchitect ECS to simplify sustaining engineering and automate operations

Features:

- Simplify software architecture (eliminate 15 components & 750K SLOC)
- Move towards disk-based archive
- Leverage new hardware technology (e.g., commodity-based systems; shared storage) to reduce hardware maintenance costs

Benefits:

- Low risk approach based on proven data pool technology
- Increased system automation; simplified hardware/software configuration
- Reduction in operational costs at ECS DAACs
- Improved data access due to increased on-line storage and commodity disks/platforms
- Provides risk mitigation for GES DAAC and LaRC DAAC ECS phase out efforts

ASDC's Role in EOSDIS Evolution

NASA HQ Science Mission Directorate approved plan for ASDC Evolution in Fall, 2005

- Develop an advanced ingest, archive, and distribution system (ANG_) to meet broader science community needs
- Consolidate LATIS (CERES) and ECS (MISR, MOPITT, TES) functionality into ANG_
  - Provides flexibility to react to researchers' data access needs (Grid services, Web Services)
  - Researchers needs change as they discover new ways to do science
  - Led to Dr. Bruce Barkstrom's vision for commodity based distributed archive
- Increase automation of MISR and CERES processing
- Reduce operations costs by 25-35% in FY08 timeframe
- New changes and revitalization will pave the way to the future
  - Reductions in operational costs provide available funds for new Earth Science applications, new missions, potential new Earth Science Data Records for decadal studies and applications to support measurement-based data

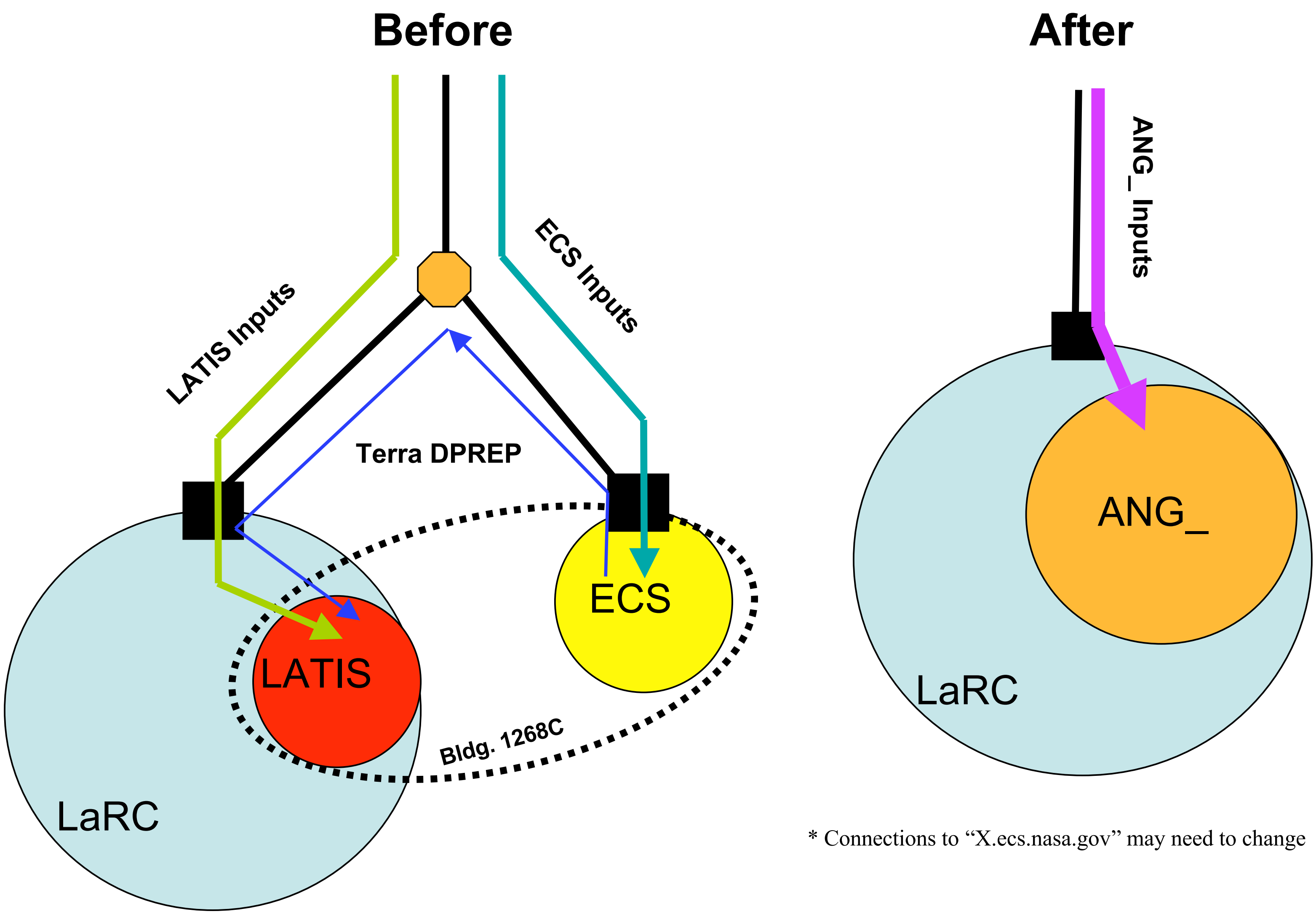
Basic Approach to ASDC Evolution

- Minimize the number of archive systems we need to manage
  - Move from two to one
  - Reduce duplication of operations
- Leverage Commodity Hardware
  - x86/PowerPC based system
  - Inexpensive Disks
- Build upon Open Source Software
  - Linux
  - JBoss – Java J2EE applications server
  - Apache
- Improve efficiency in our production environment through more automation of CERES and MISR systems

Move to a Single Archive - ANG_

- A data archive based on the Open Archive Information System (OAIS) reference model.
  - Makes it easier for ASDC to respond to changing environments
- The Name: Archive - Next Generation
  - Needed something different than LATIS/ECS, pronounceable, not already in use, and not trademarked.
  - ANG_-S: Core Archive Software
  - ANG_-L: LATIS activities running on the Core Archive Software
  - ANG_-E: Everything at the ASDC running on the Core Software

Archive Before and After



* Connections to “X.ecs.nasa.gov” may need to change

Why do we need it?

- Current archives are showing their technological age
- Use cases are expected to change into something the current architectures can't handle well
  - More and more emphasis on the broad utilization of the data
  - We will likely become input providers to general "systems"
    - Multidiscipline Research
    - Modelling Environments
  - Architecting ANG_ with the expectation of an explosion in performance requirements

ASDC Before and After

Before	After
Two completely independent systems - ECS and LATIS	One system - ANG_
Operator intensive CERES processing	Automated CERES processing
SGI MIPS / Sun SPARC focus	Commodity based - x86 (AMD/Intel), PowerPC

Evolution Effects on CERES

- ASDC is not imposing changes to external data providers
- Collection 4 and 5 MODIS data delivered from MODAPS instead of GES DAAC
- Ordering tools (e.g. Java and HTML tool) will still be available
- Subsetting will still be available
- New capabilities that Science Team can Utilize
  - Subscription
  - Data Pool

Current Status of ASDC/CERES Evolution Activities

- ASDC and CERES Science/DMT collaborating
  - CERES Automation System
  - Moving CERES Processing to commodity Systems
  - SCF and ASDC are working together to leverage our total resources

Evolution Schedule

11/15/06	ANG_-L Operations Readiness Review
11/27/06-12/28/06	LATIS to ANG_-L Mission Critical Functions Transition Period
11/27/06-2/28/07	LATIS to ANG_-L Mission Essential Transition Period
11/19/07-12/31/07	ECS to ANG_-E Transition Period
12/31/2007	ECS Roll-Off
1/2006-4/2008	CERES Automated Processing Complete

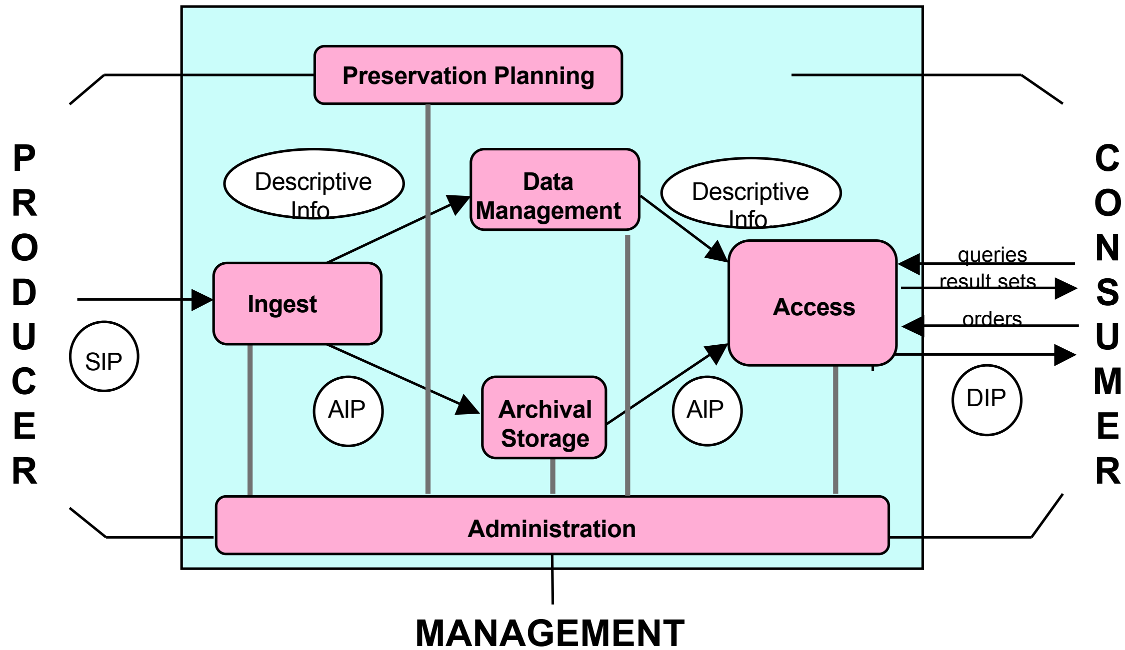
Next Steps

- Increase collaborations between CERES Science Team and ASDC
  - Take advantage of ASDC's evolved data system
  - Support New Data Fusion Initiatives
  - Joint Proposals between Science Team and ASDC to leverage ASDC Holdings

Questions?

BACKUP

OAIS Reference Model



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

The Open Archival Information System (OAIS) Reference Model is an ISO standard for describing the services an archive system must perform. ANG_ development is organized by each of the primary OAIS model component.

EOSDIS Evolution 2015 Vision Tenets

Vision Tenet	Vision 2015 Goals
Archive Management	▪ NASA will ensure safe stewardship of the data through its lifetime. ▪ The EOS archive holdings are regularly peer reviewed for scientific merit.
EOS Data Interoperability	▪ Multiple data and metadata streams can be seamlessly combined. ▪ Research and value added communities use EOS data interoperably with other relevant data and systems. ▪ Processing and data are mobile.
Future Data Access and Processing	▪ Data access latency is no longer an impediment. ▪ Physical location of data storage is irrelevant. ▪ Finding data is based on common search engines. ▪ Services invoked by machine-machine interfaces. ▪ Custom processing provides only the data needed, the way needed. ▪ Open interfaces and best practice standard protocols universally employed.
Data Pedigree	▪ Mechanisms to collect and preserve the pedigree of derived data products are readily available.
Cost Control	▪ Data systems evolve into components that allow a fine-grained control over cost drivers.
User Community Support	▪ Expert knowledge is readily accessible to enable researchers to understand and use the data. ▪ Community feedback directly to those responsible for a given system element.
IT Currency	▪ Access to all EOS data through services at least as rich as any contemporary science information system.

EEE Study Team Members

Name	Organization	Title
Moshe Pniel, Chair	NASA JPL	Manager, Astrophysics Formulation Program Office
Martha Maiden*	NASA HQ	Program Executive for Data Systems, Science Mission Directorate
Mary A. Esfandiari*	NASA GSFC	Deputy Program Director for EOS Operations & ESDIS Project Manager
Walt Brooks	NASA ARC	Chief, Advanced Supercomputing Division of the Information Sciences and Technology Dir.
Dr. Peter Cornillon	University of Rhode Island	Professor of Oceanography
Dr. Scott Denning	Colo. St. Univ	Assistant Professor, Atmospheric Science
Dr. Jim Frew	UC - Santa Barbara	Associate Professor, School of Environmental Science and Management
William Green	Retired (Cal Tech)	Manager of Infrared Processing and Analysis Center and Spitzer Science Center
Bernard Minster	UC - San Diego	Professor of Geophysics, Scripps Institution of Oceanography

** Ex Officio*

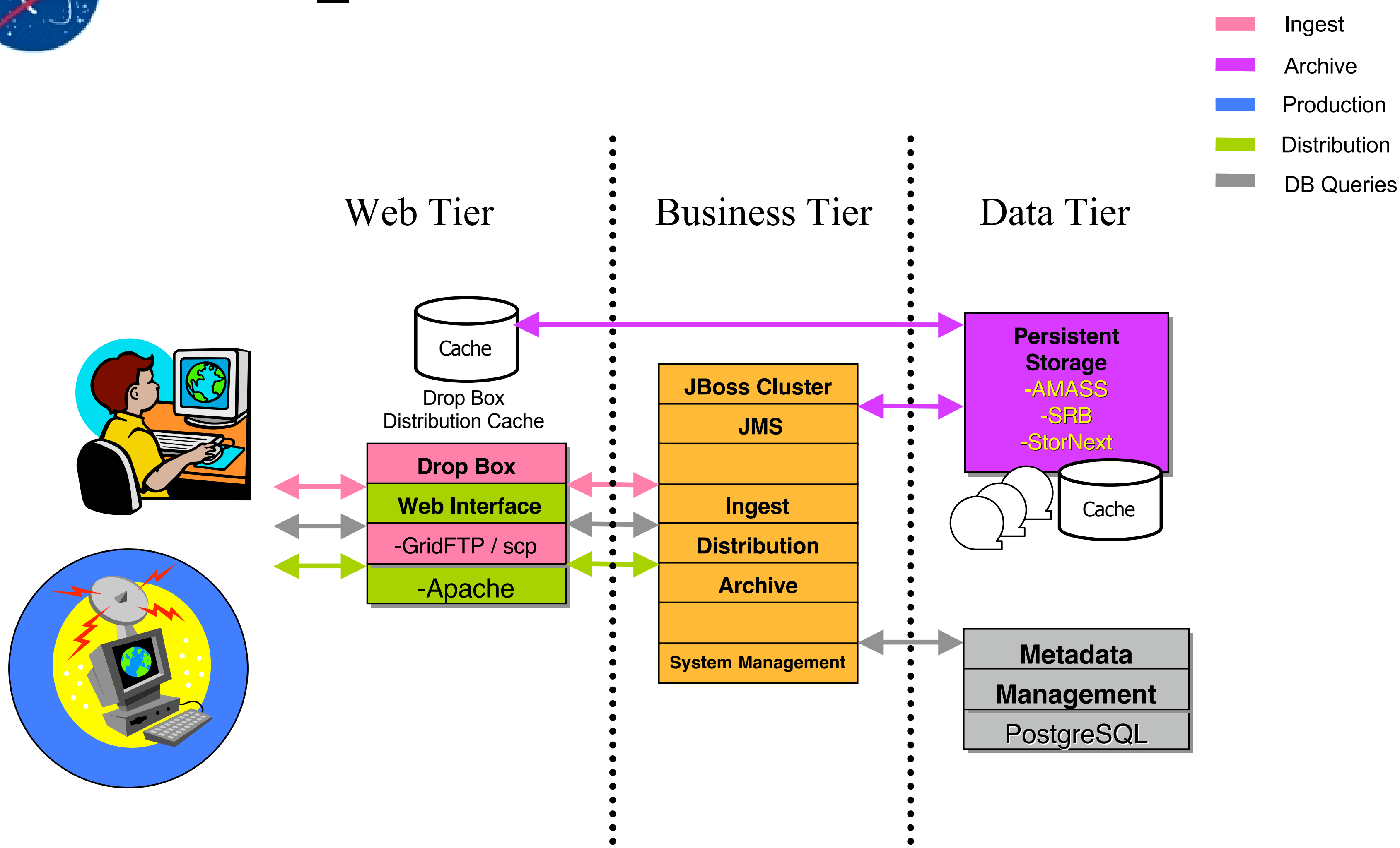
EEE Technical Team Members

Name	Organization	Title
Mary A. Esfandiari, Lead	NASA GSFC	ESDIS Project Manager & Deputy Program Director for EOS Operations
Dawn Lowe, Implementation Lead	NASA GSFC	ESDIS Deputy Project Manager
Jeanne Behnke	NASA GSFC	Acting, ESDIS Science Operations Office Manager
Chris Bock	NASA GSFC	Hardware Systems Manager, ESDIS SSDO
Michelle Ferebee	NASA LaRC	Langley Assistant DAAC Manager
Kathy Fontaine	NASA GSFC	Global Change Data Center
Michael Goodman	NASA MSFC	Earth and Planetary Science Branch
Pat Liggett	NASA JPL	Manager, Physical Oceanography DAAC
Ken McDonald	NASA GSFC	External Development & Interfaces, Long Term Archive
Ed Masuoka	NASA GSFC	MODAPS SIPS Lead
Dan Marinelli	NASA GSFC	Deputy Manager, ESDIS SSDO
Karen Moe	NASA GSFC	Earth-Sun System Technology Office
Robin Pfister	NASA GSFC	ECHO, Middleware Lead, ESDIS Project
Dr. H. Ramapriyan	NASA GSFC	Assistant Project Manager, ESDIS
Dr. Skip Reber	NASA GSFC	ESDIS Project Scientist
Curt Schroeder	NASA GSFC	EMD COTR
Ed Sofinowski	SGT, Inc.	Process & Technical Consultant
Dr. Scott Turner	Aerospace Corp.	Lead for Cost Assessment, Alternatives Analysis
Bruce Vollmer	NASA GSFC	GES DAAC Engineer

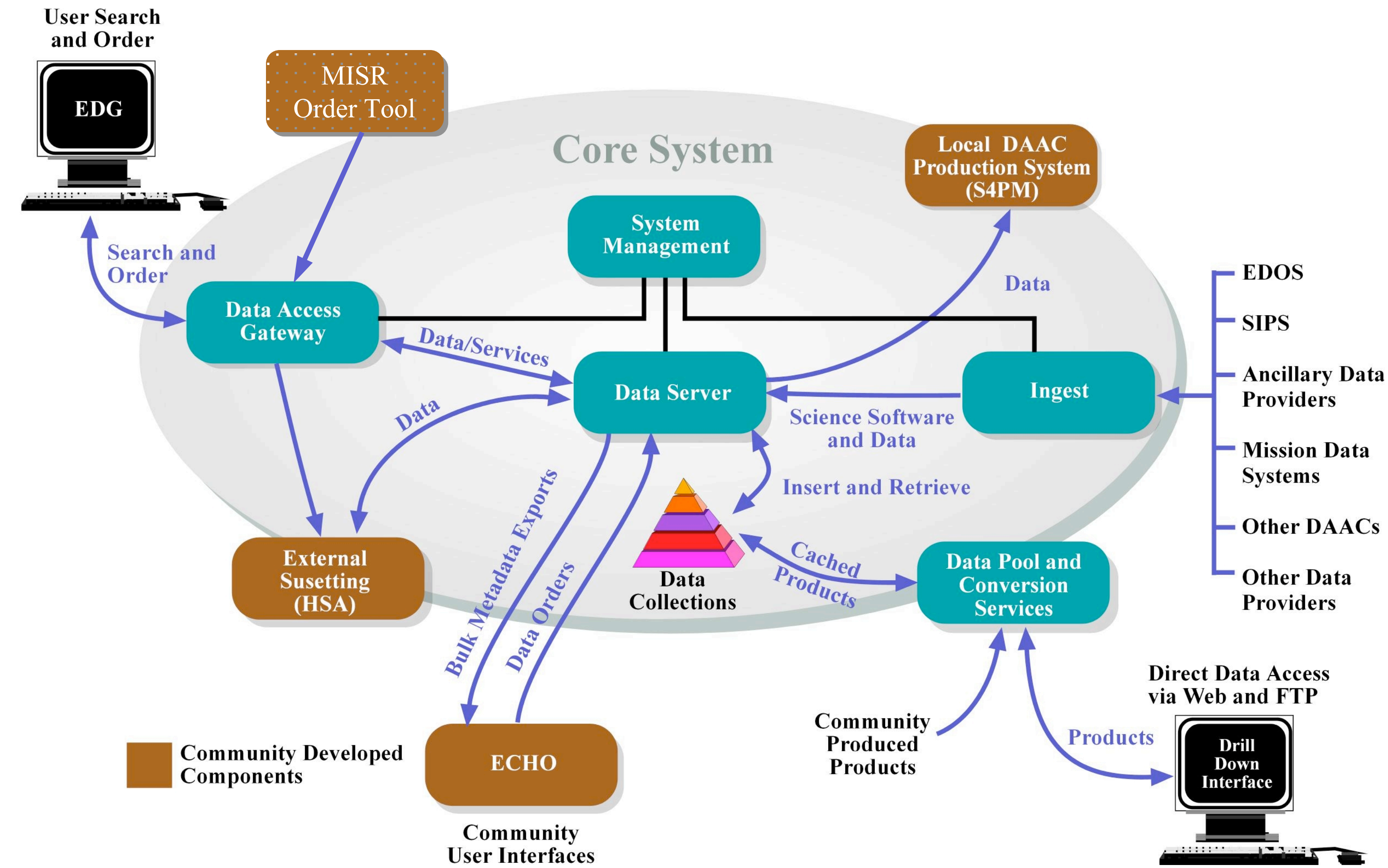
EEE Technical Team Consultants

Name	Organization	Title
Dr. Gene Feldman	NASA GSFC	Manager, OCDPS
Dr. Chris Lynnes	NASA GSFC	GES DAAC Lead, System Engineering
Dr. Erich Stocker	NASA GSFC	Manager, TSDIS/Precipitation Processing System
Dr. Victor Zlotnicki	NASA JPL	Program Manager for Climate Variability

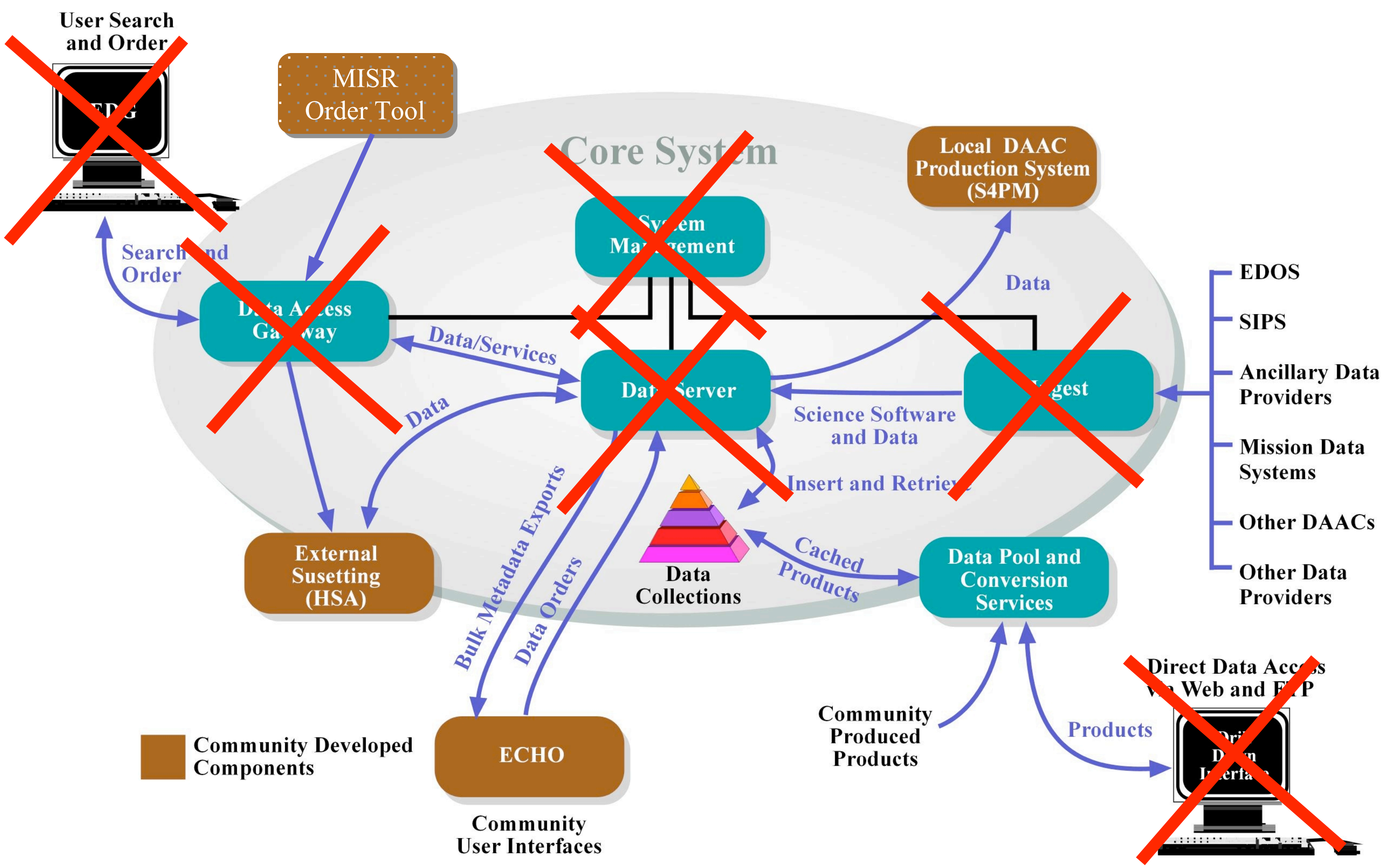
ANG_ Software Architecture/Data Flow



ECS Architecture (Before)

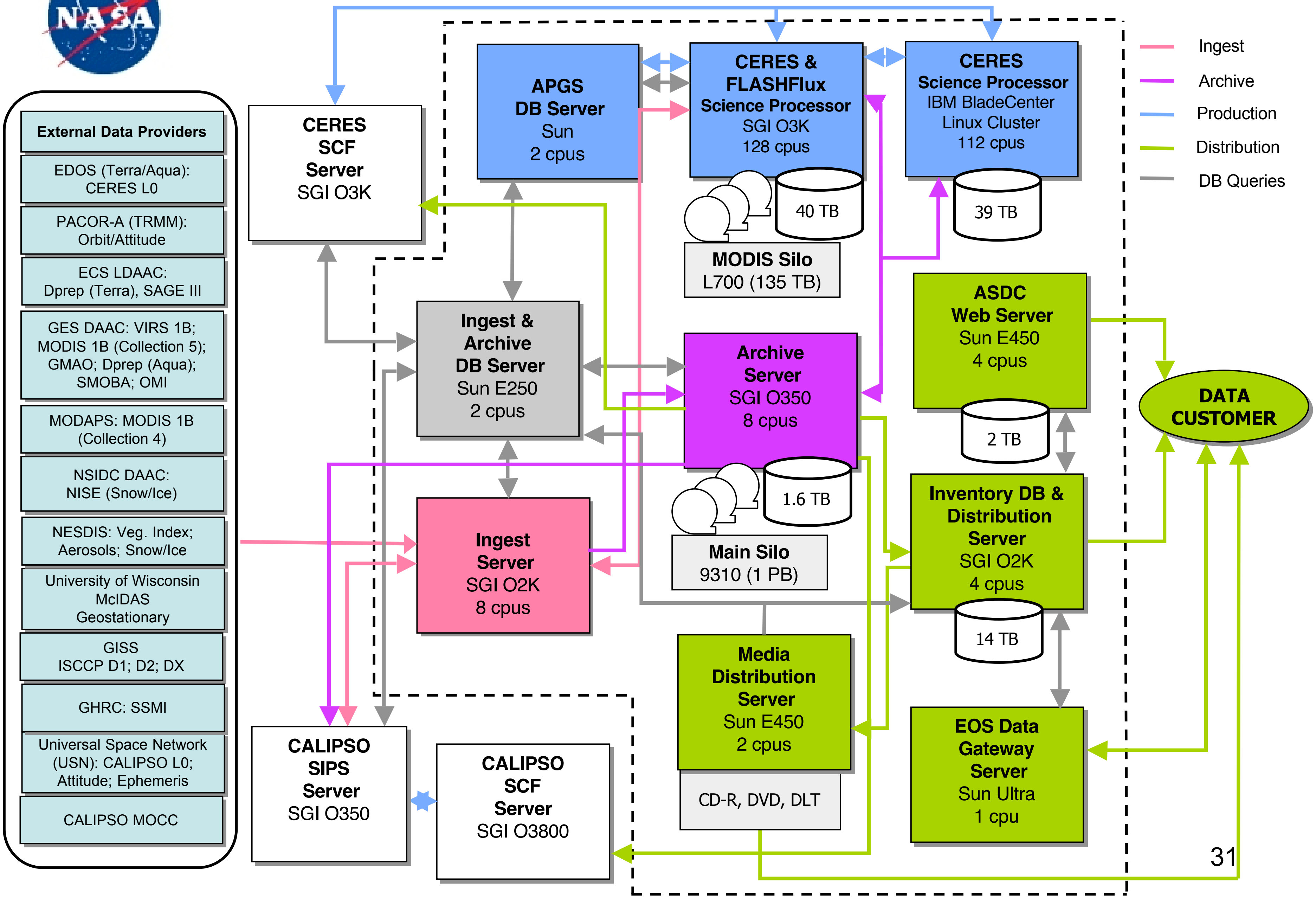


ECS Architecture (After)



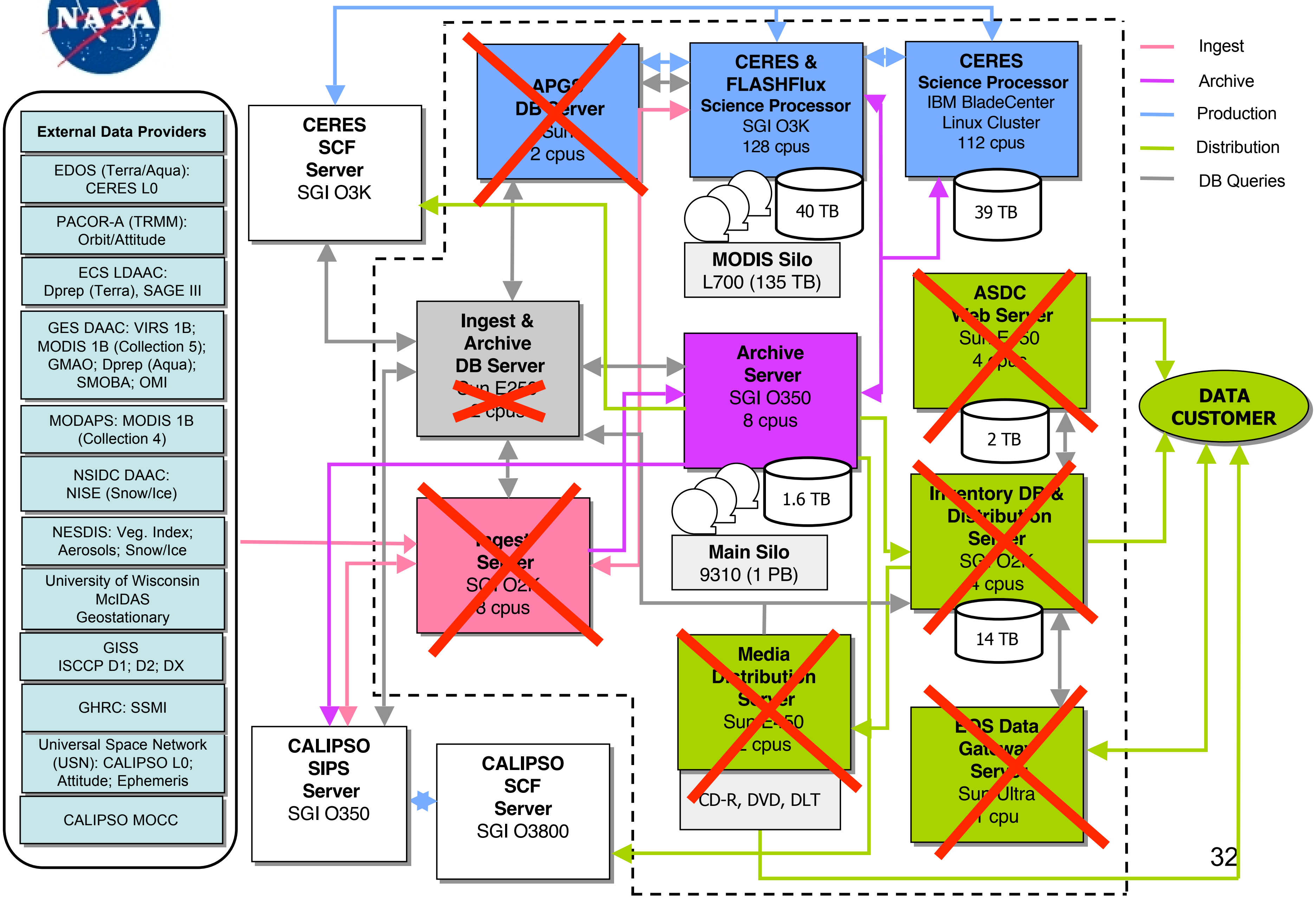
LATIS Architecture: Hardware/Data Flow (Before)

(Last Updated 03/07/2006)

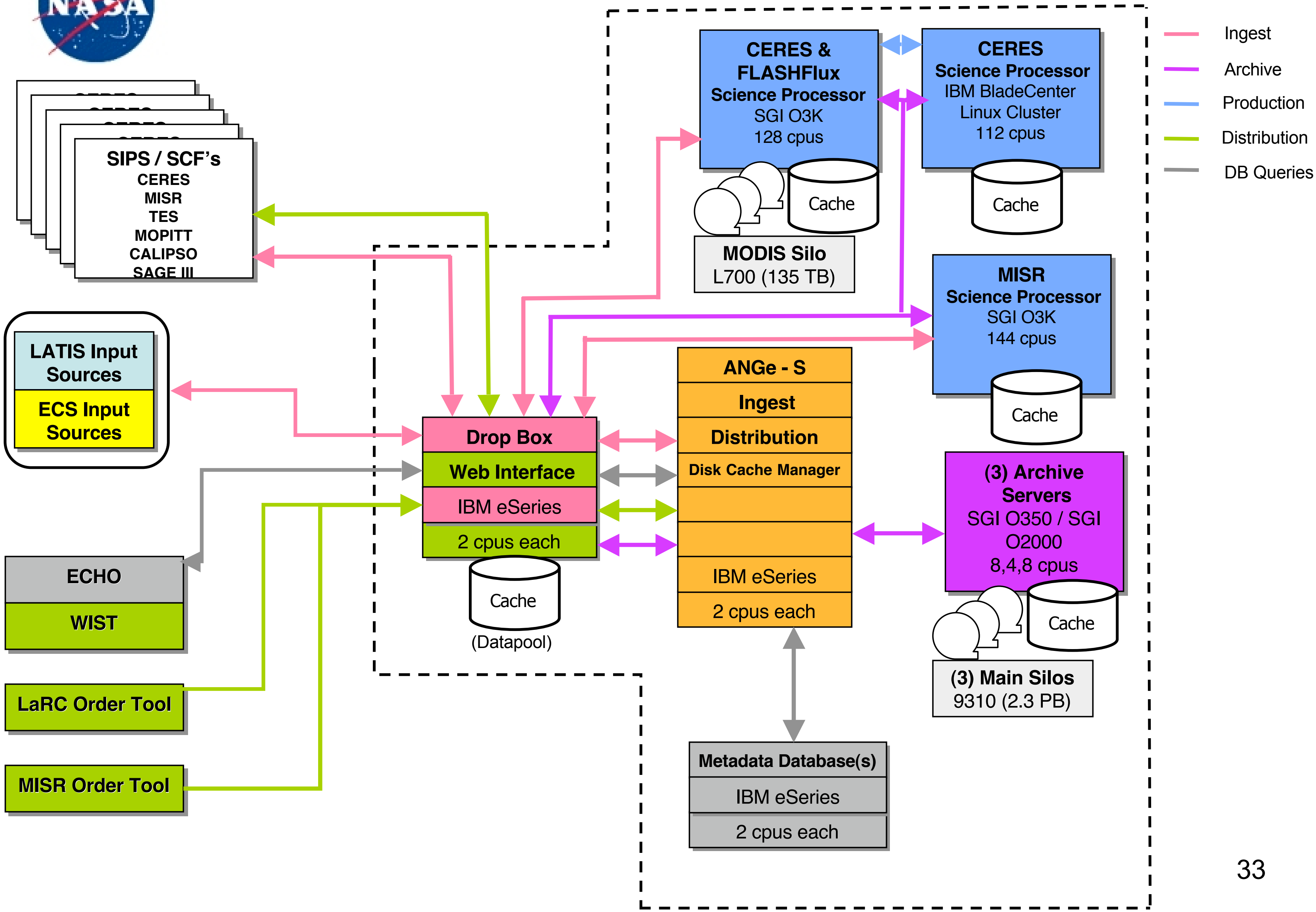


LATIS Architecture: Hardware/Data Flow (After)

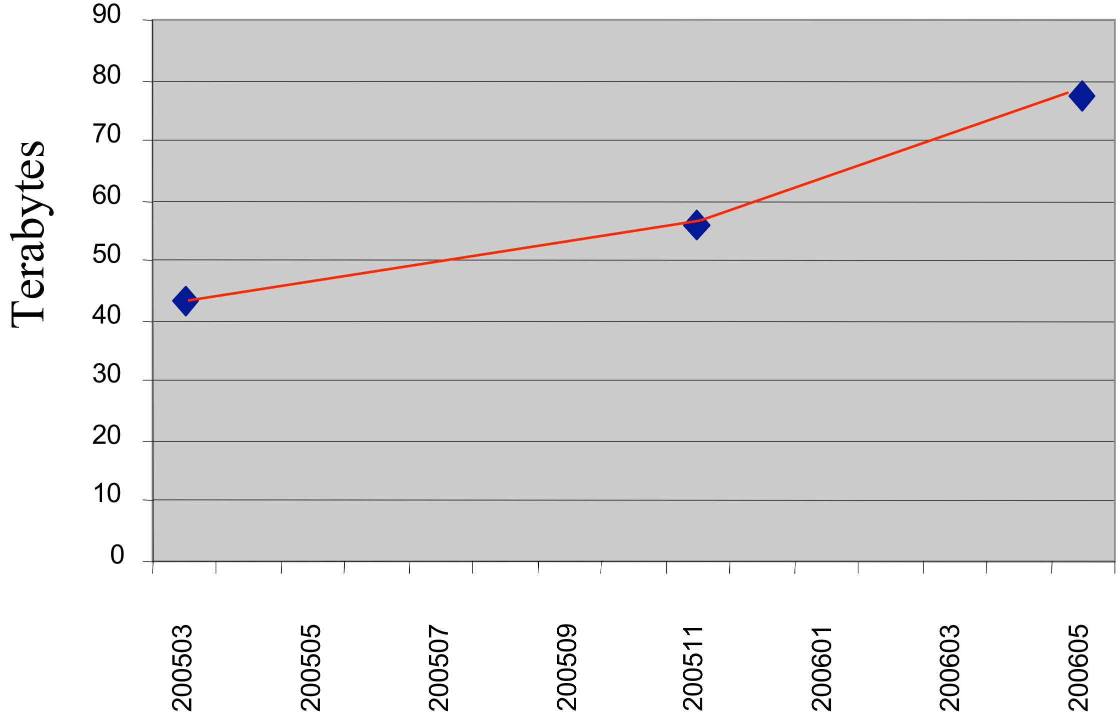
(Last Updated 03/07/2006)



Hardware After (Conceptual Architecture)



Total Data Distributed since Jan 2003

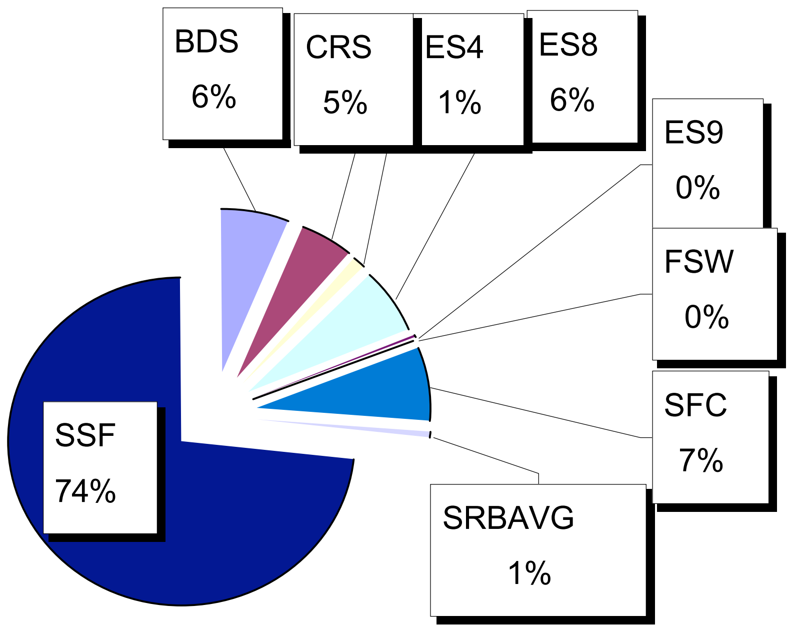


April 2006 (+5months)
Total # of Granules = 751,638
Total Volume = 79,168 GB (+22 GB)

Nov 2005 (+8months)
Total # of Granules = 593,748
Total Volume = 57,115 GB (+13 GB)

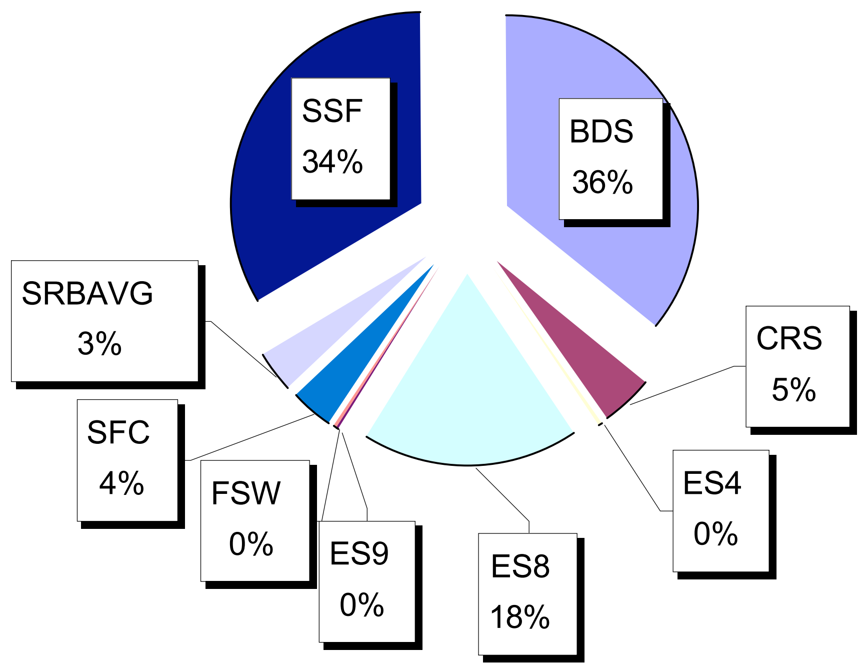
Jan 2003 to March 2005
Total # of Granules = 467,558
Total Volume = 44,093 GB

of Granules by Data Product



Nov 2005 to April 2006
Total # of Granules = 157,890
Total Volume = 22,053 GB

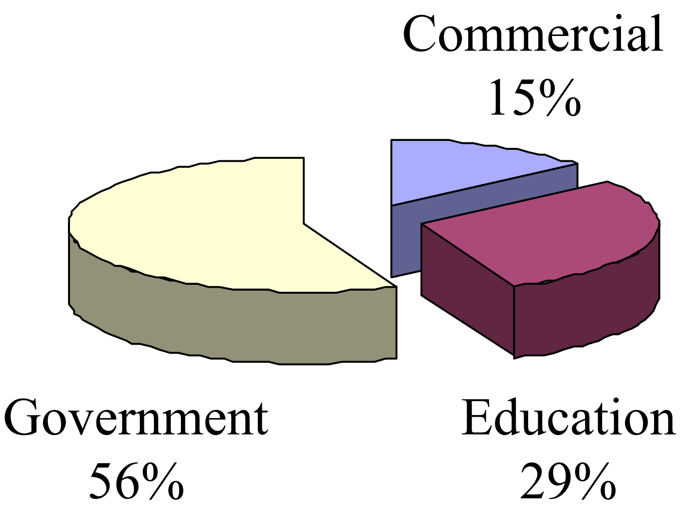
Volume by Data Product



5/3/2006

Distribution by Agency
Jan 2005 – April 2006

Data Volume



of Granules